

Solana Twitter NLP-Analysis

Twitter Topic Modeling for Crypto Hedge
Fund Research

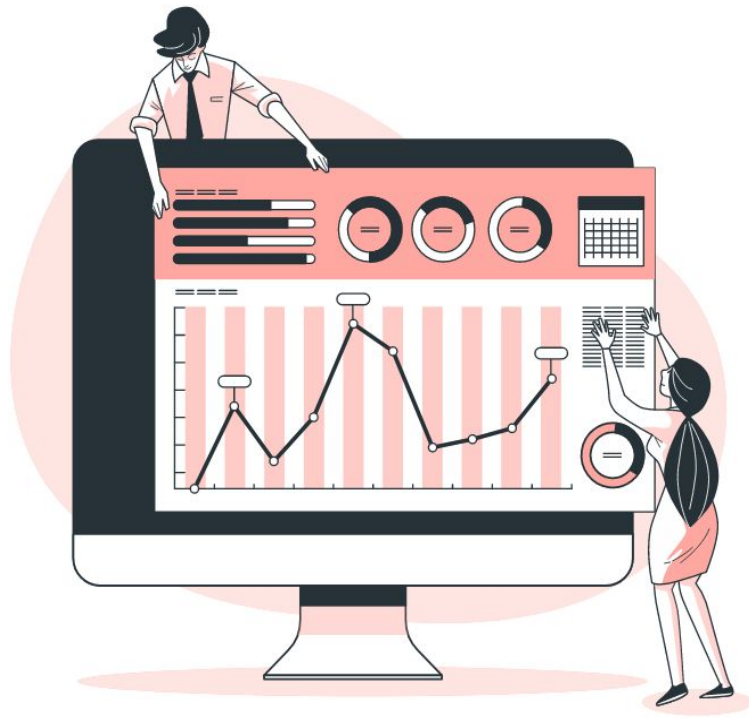


Table of contents

01 Purpose

Why?

03 Analysis Results

What's the outcome?



02 Analysis Breakdown

How was it done?

04 Future Work

What's next?

01



Purpose

Introduction



- Crypto Hedge Funds are benefitting from Crypto Boom
 - Median return of **128%** in 2020 (vs. +30% in 2019)
- These funds are using various strategies to trade crypto assets, some proving more successful than others

2020 Median Results per Strategy

294%
Discretionary Long
Only

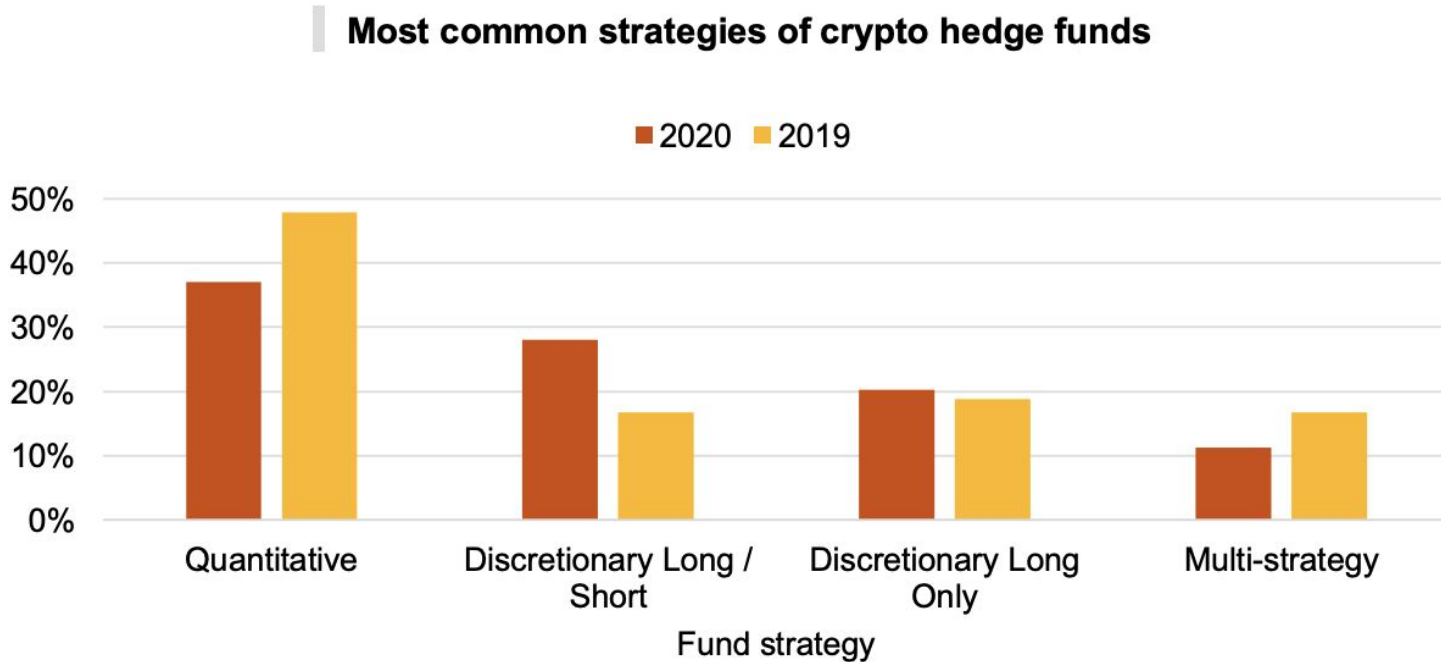
114%
Multi-Strategy



129%
Discretionary
Long-Short

72%
Quantitative

2019-20 Strategies Shows Disconnect



Goals

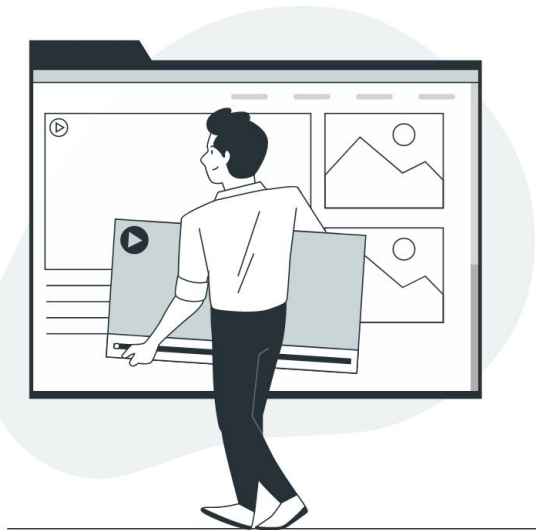


Project Goal

Empower discretionary crypto research, specifically retail investor analysis

Analysis Goal

Build an NLP topic model that can group and sort tweets effectively



02

Analysis Breakdown

Data Acquisition and Cleaning



Raw Data

Scraped 500K+ #Solana tweets using snsrape



Pre-Processing

Filtered for tweets with 3+ likes, lower-case, emojis to text, etc...

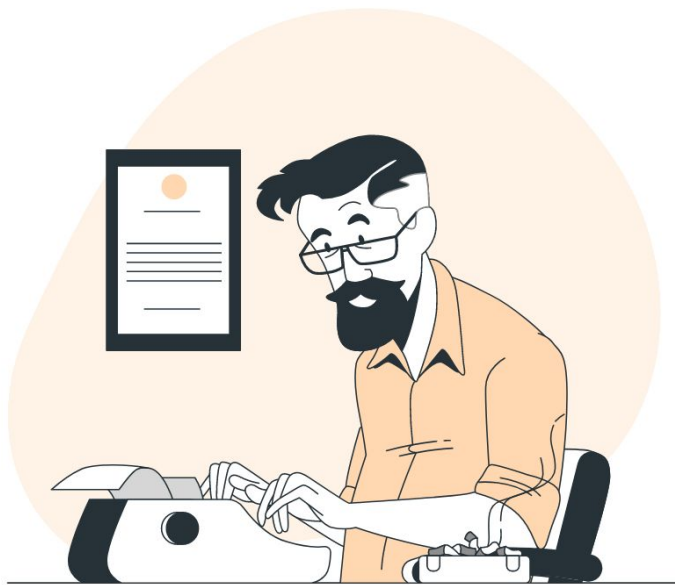


Tokenization

Tokenized tweets using NLTK



Project Workflow

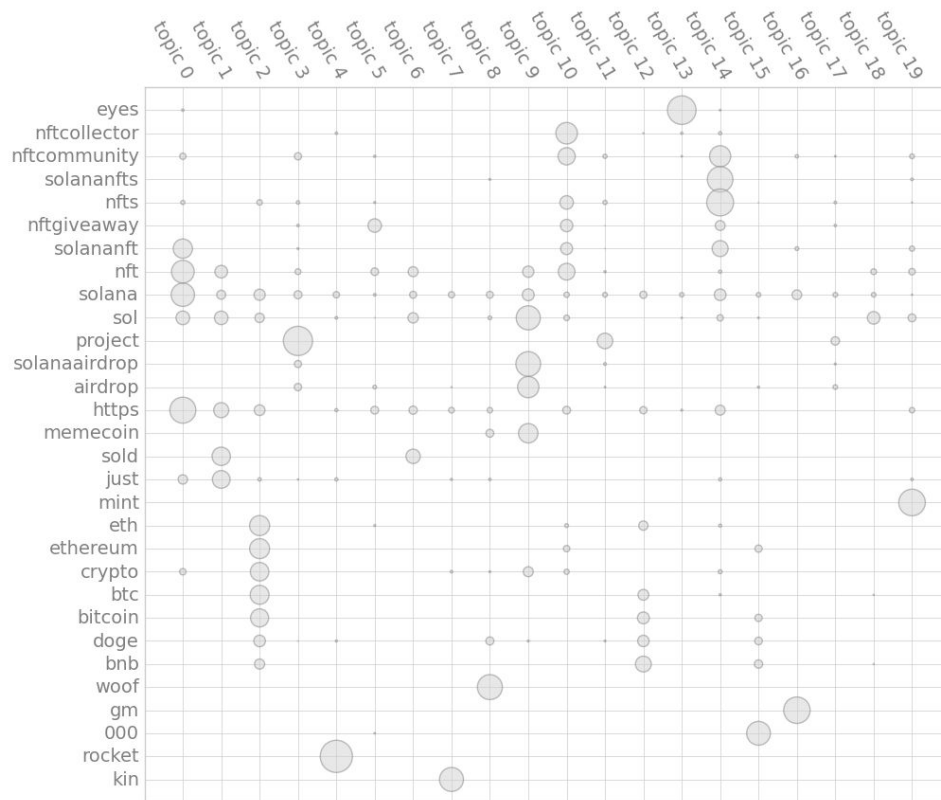


- TF-IDF Vectorizer
- NMF Modeling (20 topics)
- Textacy & Seaborn for Visualizations

03



Analysis Results

[illegible]

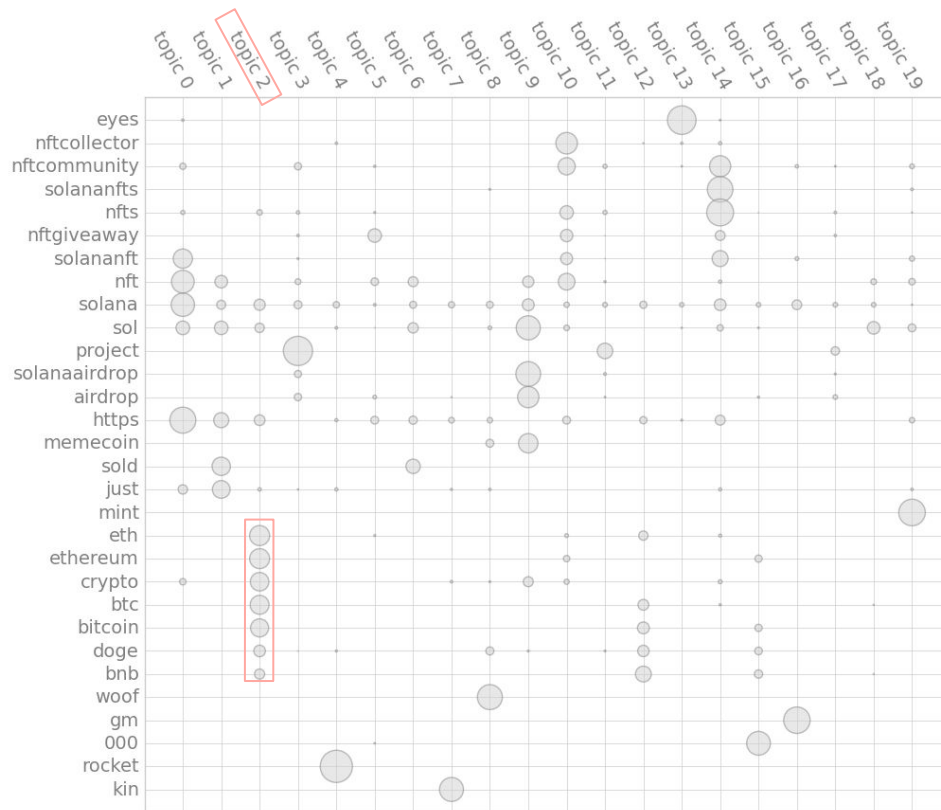
Topic 2 - Coin Comparison

“#solana could be the next
ETH or BTC ”

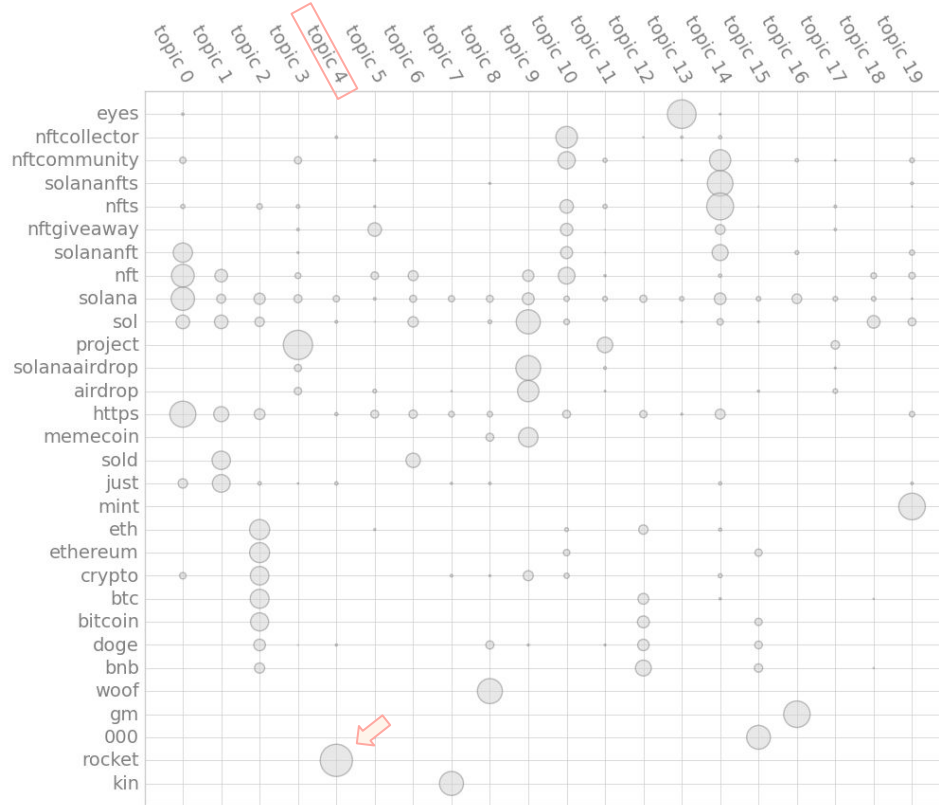
“Which one will happen
first?”

“Top 10 #Cryptocurrencies In November
2021

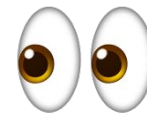
1. #Bitcoin (BTC)
2. #Ethereum (ETH)
3. #Binance Coin (BNB)
4. #Tether (USDT)
5. #Cardano (ADA)
6. #Solana (SOL)
7. #XRP (XRP)
8. #Polkadot (DOT)
9. #shibainu (SHIB)
10. #Dogecoin (DOGE)”



“\$240 #solana sell wall is
GONE 🚀🚀🚀🚀🚀🚀🚀🚀
🚀🚀🚀🚀🚀🚀 ”



Topic 13 - Anticipation

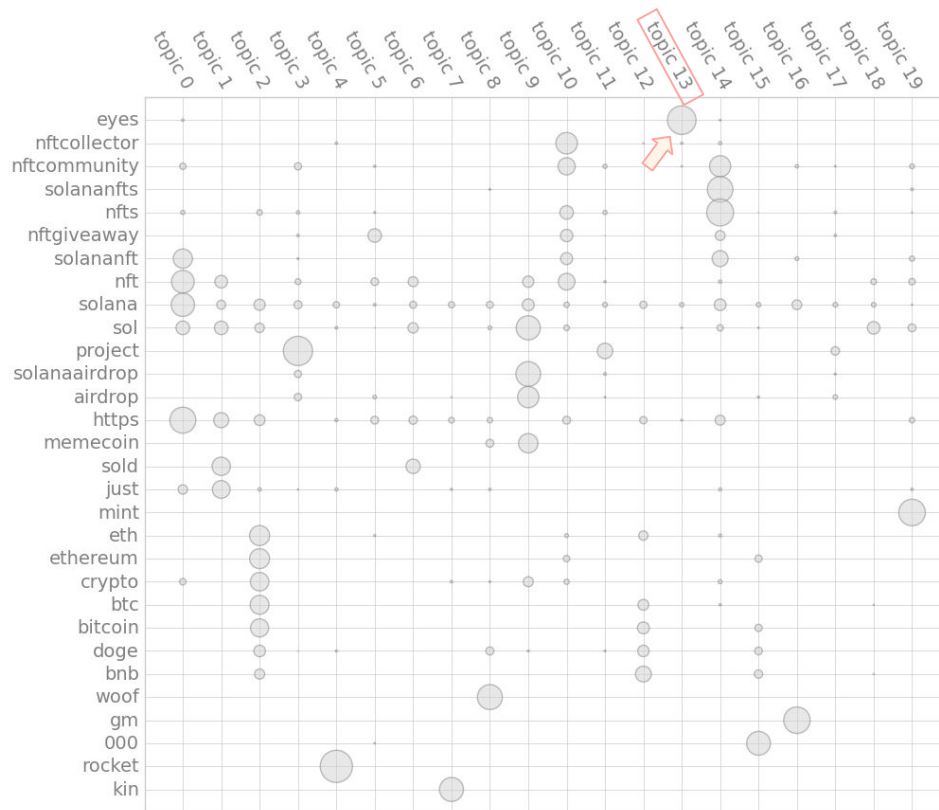


“Wowzers \$SOL #solana is moving up 📈”

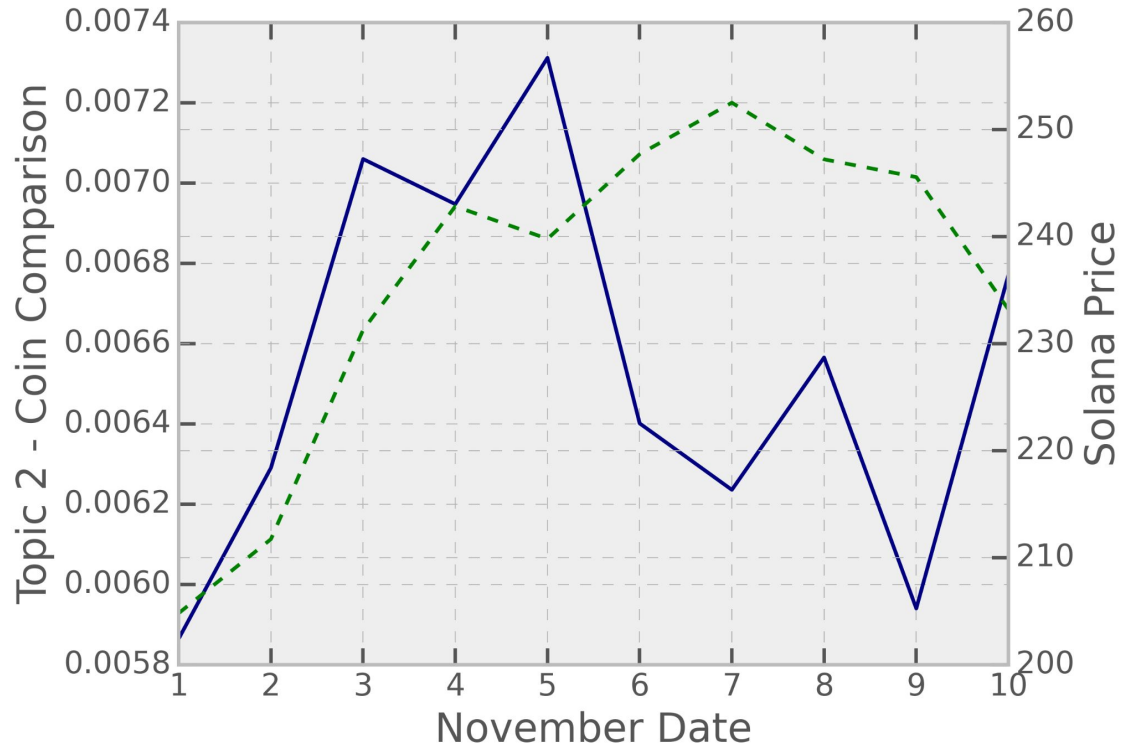


“Bullish #SOLANA 🥰”

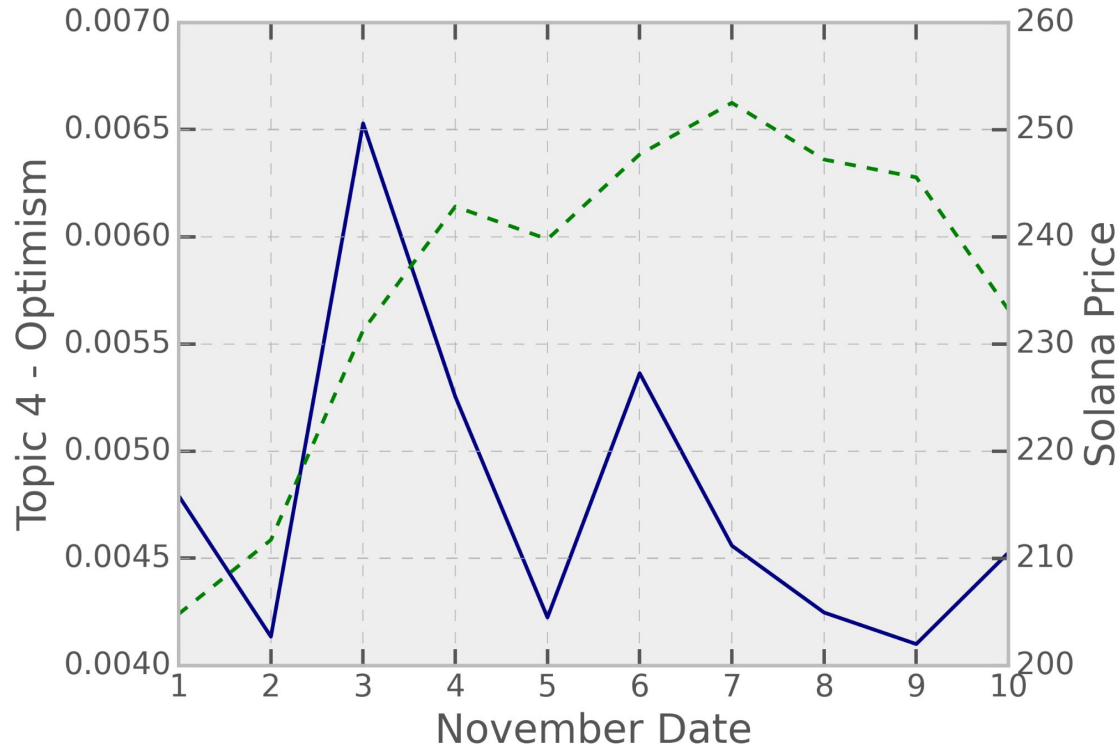
“What’s the best #Solana project? 👁️ ”



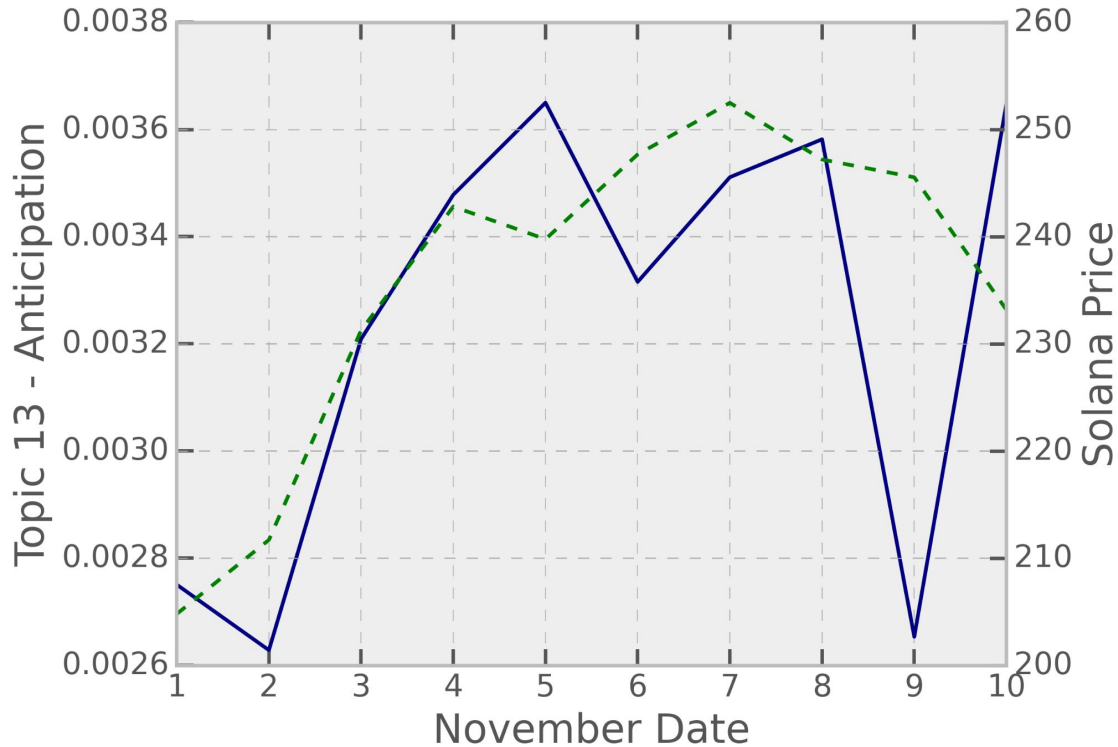
Topic 2 - Coin Comparison vs. Price



Topic 4 - Optimism vs. Price



Topic 13 - Anticipation vs. Price

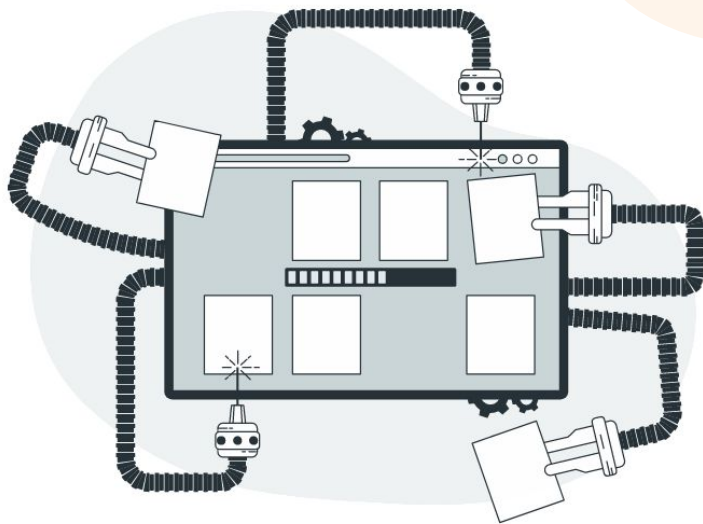




04

Future Work

Future Work



Supervised Learning

Using these outputs for a time-series linear regression problem

More Crypto Assets

Wider variety of assets, including more sentiment based and more institutional based

Sentiment Analysis

VADER, as oppose to topic modeling

**Thank
You!**

